HOW ARE NON-INFERIORITY MARGINS SELECTED IN NON-INFERIORITY TRIALS AND HOW DO THEY VARY WITHIN AND ACROSS FOUR MAJOR DISEASE DOMAINS? – A REVIEW

by
Ling Yang

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland
March, 2017

© 2017 Ling Yang
All Rights Reserved

Abstract

Objective: to summarize approaches to the selection of non-inferiority (NI) margins in non-inferiority trials and to explore the extent of variation in the selection of NI margins; **Methods:** Non-inferiority trials in stroke, cardiovascular disease, infectious disease and diabetes were searched and screened for inclusion on ClinicalTrials.gov. The primary outcome of interest was the non-inferiority margin used by trials. Potential factors regarding the selection of the NI margins were collected from full-text review. Descriptive statistics (counts and proportions) were used to report the patterns in NI margins. When appropriate, NI margins were transformed in order to evaluate and compare them. **Results:** In 8 stroke trials (42.1%), time-to-event outcome variables and NI margin expressed as a hazard ratio were generally used. The NI margins of hazard ratio ranged between 1.05 and 2. In 8 studies on coronary artery disease, all utilized in-stent late lumen loss as their primary endpoint, with NI margins ranging from 0.11mm to 0.32mm. In 8 NI trials on influenza vaccine, geometric mean titers (GMTs) and seroconversion rate were consistently used to evaluate the efficacy of vaccine. GMTs ratio of 1.5 and seroconversion rates of 10% difference were selected as the NI margins in all trials. Of 19 NI trials in diabetes, most (68.4%) used the change in HbA1c from baseline as the primary outcome and absolute difference between the changes as the comparison statistic. NI margins ranged from 0.3% to 0.5%, but 7 of 13 used a margin of 0.4%. Pooling all 47 trials across these 4 different disease areas, we found a mean of 1.43 and a 95% two-sided confidence interval from 1.33 to 1.54 on the transformed relative scale for the NI margins. **Conclusion:** There is no evident or consistent pattern on the selection of NI margins and the variation of NI margin selection is large both within specific disease area and across different disease areas.

Table of Contents

Introduction 1

Methods 7

- **Literature Searching** 7
- **Data Extraction** 9
- **Statistical Analyses** 10

Results 11

- **Stroke Trials** 13
- **Cardiovascular Trials** 15
- **Pneumonia/Influenza Trials** 16
- **Diabetes Trials** 16
- **Overall** 17
- **Rationales for the Selection of NI Margins** 19

Discussion 19

- **Limitations** 25

Conclusion 26

Bibliography 27

Tables 31

- **Table 1. Basic characteristics of the included non-inferiority trials (N=74)** 31
- **Table 2. Design of the included non-inferiority trials and their non-inferiority margins** 32
- **Table 3. Non-inferiority margins on relative scales by disease domain** 36
- **Table 4. Rationales for the selection of non-inferiority margins in the included NI trials** 37

Figures 38

- **Figure 1: Non-inferiority margins on relative scales for included NI trials by disease domain** 38

CV - Ling Yang 39

Introduction

Randomized clinical trials (RCTs) are deemed to be the gold standard in investigating the efficacy or safety of a new treatment, and there are three major types of RCTs based on their purposes: 1) superiority trials, which aim to demonstrate that a tested treatment is better than a control; 2) equivalence trials, which aim to demonstrate that a tested treatment is similar to an existing treatment within a specified margin; and 3) non-inferiority (NI) trials, which aim to demonstrate that a new treatment is not worse than the existing treatment within a pre-specified margin. NI designs are useful in situations where the efficacy of a tested drug is thought to be roughly the same as the comparator and the benefits of comparator are known, but the tested drug has additional advantages, such as fewer adverse events, easier use, or reduced cost [1]. NI designs can also be used to indirectly show the efficacy of a tested drug over a placebo when the use of a placebo is unethical such as when an efficacious standard treatment is already available and it would be improper for patients in such a trial to be given an inferior treatment [2]. Likewise, NI trials can test pharmacologically related compounds to see if they are similarly effective [3]. Because of their many strengths, non-inferiority randomized trials are gaining in popularity in recent years, especially in the fields of oncology, infectious disease, or cardiovascular disease [4]. However, non-inferiority trials also face methodological challenges in determining the appropriate "non-inferiority margin" – a pre-specified amount by which the extent that the test treatment is inferior to the control is not allowed to exceed when we intend to prove that the test treatment is not worse than the control. The NI margin is thus the largest loss of effectiveness to establish non-inferiority.

When conducting a non-inferiority trial, the alternative hypothesis is that the outcome of the tested treatment may show slightly less effectiveness to the active control, but the difference between the groups is no more than a pre-determined non-inferiority margin. Margins that are chosen to be too large may lead to trials that are easier to complete but show purported non-inferiority while potentially being less effective than the comparator. General guidance exists on the determination of NI margins. The International Conference on Harmonization (ICH) has published a guideline entitled Guidance on Choice of Control Group and Related Design and Conduct Issues in Clinical Trials, which is referred to as ICH E10 [5]. The ICH E10 guideline provides some general principles for the selection of appropriate non-inferiority margins. According to the ICH E10, the selection of the non-inferiority margin should be based on both statistical reasoning and clinical judgment. The ICH E10 also points out that the selection of the non-inferiority margin should be based on historical experience in placebo-controlled trials, and be suitably conservative about the effect size of the active control derived from previous placebo-controlled trials. However, the ICH E10 only provides general principles on the selection of NI margin and there are no specific standards for the determination of NI margin, which can give rise to controversial issues in the research process, such as the possibility for the researchers to twist the NI margins to make their studies feasible, while at the same time remaining able to justify the selection of the NI margin even when it is not reasonable due to the lack of clear guidance or recommendations.

Some regulatory authorities have attempted to provide guidance regarding the choice of the NI margin in specific therapeutic areas, and researchers have conducted much exploration into the best approach. A recent U.S. Food and Drug Administration (FDA)

guidance on NI trials specifically states that determining the NI margin is the 'single greatest challenge' as it ties closely into assay sensitivity and is based on the constancy assumption regarding the effect of active control in the NI trial [6]. The constancy assumption states that the control treatment will show the same measurable efficacy versus placebo in the current trial as it did in the historical trials. The FDA guidance presents a more concise approach to the choice of margin that can be easily understood by general audiences, involving both statistical reasoning and clinical judgment, stating that NI trials are designed to show that any difference between the test drug and the active control is small enough to allow one to conclude that the test drug has at least some effect or an effect that is not too much smaller than that of the active control [7]. When designing a non-inferiority trial, the effect size of the active control compared with the placebo needs to be estimated on the basis of historical data from previous trials. This effect size is called M1. Once this number is available, clinical judgment can be used to define the non-inferiority margin (M2), the largest loss of effect relative to the active control that is clinically acceptable. Even though M2 can be as large as M1, it is more scientifically meaningful to choose an M2 that is a fraction of M1 which is also clinically acceptable. In short, the non-inferiority margin should be no larger than the threshold of clinical relevance as well as being small enough to exclude placebo.

The FDA also describes a method to select the NI margin [8]. First, the total assumed effect of the active comparator over the placebo must be estimated, and a conservative estimate of this effect should be taken to ensure that the test drug has an effect that is greater than zero. This conservative estimate can be the lower limit of the confidence interval of the difference in effect when comparing the active control with the placebo.

Although conservative, this minimum treatment effect helps to protect against an overestimation of the assumed active control's effect, leading to a poor choice of the NI margin. Second, a preserved fraction of the estimated effect should be determined demonstrating that the treatment is not unacceptably worse than its active comparator based on clinical judgement.

Several factors can affect the choice of an NI margin. One potential factor is the baseline risk in the control group of the trial. Another candidate factor that may influence the NI margin is the type of outcome. Indeed, a previous exploratory study [9] showed that the NI margin should be significantly lower when mortality was the primary outcome. Moreover, if the treatment reduces mortality, the life expectancy of the patient may also play an important role because people may be less willing to accept an increase in mortality if the patient still has many years of life left. Because both mortality and average life expectancy may vary by disease area, customary NI margins may also vary across different disease areas. Another candidate factor that may affect the selection of the NI margin is the type of benefit of the test treatment. If the test treatment has other benefits that the control does not, including being much less costly, causing fewer side effects, or being less invasive, both the clinical practitioners and the patients may be willing to accept the test treatment even when its efficacy is not as large as another treatment. The loss of the effectiveness could be offset by other benefits of the test treatment that patients are more concerned about.

However, there are also invalid factors that affect the selection of the NI margins. When investigators expect that there would be recruitment difficulties, for example trials on pediatric patients, or the budgets would be tight due to insufficient funding, they may

be willing to accept a larger loss of efficacy and thus a larger NI margin. A larger NI margin implies a smaller sample size, and a larger margin may make a trial more feasible with limited resources. A larger NI margin also makes the NI trial less conservative by tending to produce false positive results, saying that the test treatment is not inferior to the control while the test treatment is in fact inferior to the control. Meanwhile, small sample size would make the study under-powered. The detectable difference between the two treatment groups increases, and a larger NI margin is then selected to ensure that the difference between the two groups could still be captured and detected, thus making it easier to get the conclusion that the test treatment is not inferior to the control. These are not the correct ways to select the NI margins, but these factors generally affect the process of the selection of the NI margins often.

There is little uniformity on how to determine the NI margin when designing or interpreting a non-inferiority trial. Some researchers would select the NI margin based on the effect of the control treatment against placebo or even on the lower confidence limit of the effect's estimate. But this approach can lead to conclusions that are difficult to justify; for example, the more effective the standard treatment, the larger the loss of effectiveness that would be accepted in a non-inferiority trial simply to remain better than placebo. Others might select the NI margin depending on the severity of the primary endpoint, with a smaller NI margin when mortality, disability or other serious events of interest are used. The FDA has also proposed a loss of effectiveness of 10% in absolute terms as compatible with non-inferiority for anti-infectious or antiretroviral therapies [10]. However, such a simple rule is not easily applicable to all disease areas and is not related to the advantages of the test treatment. More recently, the FDA and the European Medicines Agency (EMA)

have proposed to select the value that preserves at least 50% of the effect of standard treatment compared with the placebo [11], but this is a liberal approach and could lead to unreasonably wide NI margins. Additionally, in their guidelines for trials in certain therapeutic areas (such as diabetes mellitus and infectious diseases), the FDA and EMA provide more explicit guidance on how to design a non-inferiority trial. What is surprising is that there are discrepancies between the FDA and EMA in the guidelines that recommend a specific non-inferiority margin. For example, in the 2008 draft FDA guidance for diabetes mellitus, a non-inferiority margin for HbA1C reduction is suggested to be 0.3% or 0.4%, while the 2011 EMA guideline suggests a non-inferiority margin of 0.3% [12]. A difference of 0.1% of HbA1C may not be clinically meaningful or useful, but this difference will result in totally different sample size calculations across non-inferiority trials. In practice, the non-inferiority margin, is often "negotiated" between the sponsor of the trial and the regulatory agencies. This happens because the threshold of clinical relevance is very difficult to determine, and as one European regulatory assessor pointed out during a scientific advice meeting, "a smaller margin would have been better" [13], which is of course true for any non-inferiority margin.

Although the selection of the NI margin is complicated, its choice influences the outcome interpretation of the NI trial. Setting an inappropriate margin may result in inappropriate conclusions. In NI trials, a test treatment may be approved for marketing due to the significant benefits, such as it being less toxic, being easier to administer, or being less expensive, even if it is less effective than the standard control. A non-inferiority margin that is too wide may jeopardize the results as well as diminish the power of the study by producing a false positive result and saying that the test drug is not inferior to the control

while in fact it is inferior to the control, and encourage acceptance and use of less effective therapies, thus resulting in serious injuries to patients. Interpreting the results of an NI trial requires an assessment of the rationale for the design and the assumptions underlying the choice of the non-inferiority margin. Because the choice of the margin is to some extent arbitrary, researchers may be tempted to redefine the margin once the results are in, to claim non-inferiority. The margin must be prospectively defined at the start of the study using both statistical methods and clinical judgement. Previously it was found [14] that in 22% of the non-inferiority trials examined, the choice of the non-inferiority margin was based merely on the judgements made by the investigators, such as margins that could cause clinical difference, or selections that were used in previous NI trials.

In many situations, the NI margin is simply stated but not justified. A previous study [15] found that NI margins vary between medical specialties, but the reasons for this phenomenon are unclear. As a result, little is known about the reasoning that researchers use in selecting the NI margin. Furthermore, whether researcher-selected NI margins reflect patients' priorities is unclear. In this review, we will summarize the selection of non-inferiority margins in general NI trials and explore the extent of variation in NI margin selection, in the hope of obtaining the underlying pattern of NI margin selection and giving some applicable recommendations to guide the determination of NI margins when designing non-inferiority trials.

Methods

Literature Searching

To identify potential non-inferiority trials for analysis, I pre-specified several disease domains where non-inferiority trials are likely to be conducted. The areas of stroke

and cardiovascular disease were chosen as representatives of circulatory diseases, where patients' conditions are generally serious or acute, and superiority trials with placebo control are unlikely to be considered. Meanwhile, superiority trials with active control in both areas are usually not the primary goal of investigating test treatment when the actual purpose is to prove that the test treatment is equivalent or at least not worse than the existing treatment while perhaps having other benefits. Diabetes was chosen as a category of endocrine disease as well as chronic disease, and with the reason that diabetes has long been a global public health concern [16]. More than 415 million people are living with diabetes and 5 million deaths are attributable to diabetes globally [17]. New treatment methods are continuously brought up and evaluated using clinical trials, especially non-inferiority trials when most test treatments have similar efficacy to existing treatments but are less costly to reduce the health expenditure due to diabetes for both patients and society. Pneumonia and influenza were chosen as categories of infectious and acute diseases, where the non-inferiority margins are relatively more well-established. By further reviewing and investigating the NI margins that researchers used in published trials, I can examine the extent to which the recommendations are in fact followed in trials, in the hope of standardizing the selection of NI margins for infectious diseases and bringing up with some consistent guidelines in the selection of NI margins that researchers could follow.

Records of non-inferiority trials were searched on ClinicalTrials.gov, with the search terms being "non-inferior" or "noninferior" or "non-inferiority" or "noninferiority". Four independent searches were done, one each the four different disease domains described above. For each search, the basic search terms to select NI trials were retained, and the "Conditions" area in "Advanced Search - Targeted Search" was additionally

specified as "stroke" for the stroke domain, "cardiovascular or coronary" for the cardiovascular disease domain, "diabetes" for the diabetes domain, and "pneumonia or influenza or flu" for the infectious disease domain. The records were screened on ClinicalTrials.gov for eligibility. Trials were included if they met all of the following criteria: the trial was completed by the time I conducted my search, the trial design was non-inferiority, the trial focused on one of the four disease domains, and corresponding publications were attached. The first 20 records in each domain with identified publications either for the protocol or the final report were included, in consideration of the limited time and resources I have for this study. The publications were then retrieved and downloaded from PubMed or the source journal for future review.

Data Extraction

The primary outcome of interest for my work was the non-inferiority margin used by the trials. During the full text review process, the following information was also collected as factors potentially related to the selection of the NI margin in each non-inferiority trial: disease area, whether the study was randomized, phase of the trial, blinding/masking status, test treatment and control, number of treatment arms, whether the study was placebo controlled, country or region the study was conducted, trial duration, participants' age range, sample size, primary and secondary endpoints, primary analysis method and significance level, comparison statistics (measures of association), absolute effect sizes of both the test treatment and control, and the rationales on how the trial selected its NI margin. The collected information and variables were then aggregated and input into an Excel table for further analyses.

Statistical Analyses

Descriptive statistics were used to report the count as well as the proportion of non-inferiority trials with specified characteristics. The NI margins collected from different trials were often reported in different scales – absolute scale, relative scale, and hazard ratio scale. The absolute scale means that the NI margin is expressed as the absolute difference between the observed values of the two treatment groups, and the unit of the NI margin is the same as the unit of the outcomes. For example, in a study of COPD (chronic obstructive pulmonary disease), an outcome might be the observed value of trough forced expiratory volume in 1 second (FEV1) at the end of a treatment period and the values usually lie between 1000ml to 1500ml. A NI margin of 50ml to evaluate the difference between the groups would be in an absolute scale. The relative scale means that the NI margin is expressed as a ratio comparing the two treatment groups. Using the above COPD study as an example, a NI margin of 0.8 comparing the value of FEV1 in the test treatment group to that in the control group would be on a relative scale. The hazard ratio scale means that the NI margin is expressed as a hazard ratio between the two treatment groups, with the study outcome measured as the time-to-event. To make the scales as consistent as possible and facilitate the comparisons of NI margins between studies, I transformed NI margins in absolute scales into relative scales using the effect size that the study reported for the control group. The transformation was done as follows: NI margin on relative scale = (NI margin in absolute scale + effect size of the control)/effect size of the control. While the NI margins in hazard ratio scale incorporated time, they could not be simply transformed as above, and were analyzed independently. For both scales, either relative scale or hazard ratio scale, the mean and 95% confidence interval for NI margins were calculated within

each disease domain as well as across different disease domains to evaluate the variation of NI margins. All analyses were conducted using StataSE 14.1 (Stata Corp LP, College Station, TX, USA).

Results

Among the 80 non-inferiority trials identified and reviewed, I excluded 6 trials from further study because the non-inferiority margin was not described (2 trials), or the trial was stated to be non-inferiority but the statistical analyses that the researchers finally conducted were superiority (4 trials). The remaining 74 trials were retained and their general characteristics are described in Table 1. As shown, most trials were randomized (98.6%), with just one study being an observational matched case-control study conducted in the Netherlands. Many trials had only two treatment arms (89.2%) without a placebo control (96%), presumably because trying to prove that the test treatment is not worse than placebo is the same as trying to prove that the test treatment is not worse than nothing, which would be of minimal scientific significance. The masking status of the trials varies, with nearly a third of them (32.4%) being open-label and a quarter of them (24.3%) being double-masked (both subject and investigator). The decision whether to mask or not in a trial heavily depends on the treatment arms and outcomes of interest. When the treatments' distinguishable characteristics – including route of administration, unique patients' reactions – are easily to discern, masking is not feasible. Subjective outcome measures benefit from masking to minimize information bias, while objective outcome measures may not call as heavily for masking. Almost eighty percent of the trials (79.7%) didn't report or indicate their study phase, and among those that did, most were phase III trials, meaning that the efficacy or effectiveness of the new treatment is being evaluated. 7 trials

(9.5%) were conducted on children, 53 trials (71.6%) were conducted on adults, and 14 (18.9%) of them did not have age limitations for including participants. Nearly half of the trials (46.0%) were performed multi-nationally, with multiple clinical sites across different countries or regions. The purposes of conducting multi-center trials are to enlarge sample size and include a heterogeneous population that helps assure that results will be generalizable. 36 of the 74 trials (48.6%) were analyzed using an intention-to-treat approach, in which patients are included into the analyses as randomized. In superiority trials, intention-to-treat analysis is unbiased, because it keeps the randomization intact and analyzes as the patients were randomized regardless of what the patients did. By ignoring the potential cross-overs between the test treatment and the control which would make the two treatments look more similar, the intention-to-treat analysis in superiority trials is underestimating the effect size of the test treatment compared to the control and being conservative. In non-inferiority trials, however, by ignoring the potential cross-overs which would make the two treatments look more similar, it's easier to get a conclusion that the test treatment is not inferior to or different from the control, which is a positive result for NI testing and thus is less conservative. 15 of the 74 NI trials (20.3%) were analyzed in the per-protocol population (as-treated). In a per protocol analysis, patients are analyzed as the treatment they actually receive, which may be biased but is truly conservative for non-inferiority testing and is the recommended primary analysis method for non-inferiority trials [18]. The statistical significance level for most trials (94.6%) are two-sided 95% or one-sided 97.5% with type-I error α=0.05.

Table 2 presents the NI margins and the potential correlating factors that may be associated with the selection of NI margins for the 74 NI trials. 19 trials are focused on

stroke, 20 trials are for cardiovascular diseases, 16 trials are for infectious diseases (including pneumonia as well as influenza), and 19 trials are in diabetes. Among the 74 total NI trials, 82 primary outcomes of interest were identified, due to the fact that a single trial might report multiple primary outcomes and have multiple corresponding non-inferiority evaluations. There are three types of measurements regarding primary outcomes: measured as binary (n=35, 42.7%), as continuous (n=38, 46.3%), and as time to event (n=9, 11.0%). For the 35 primary outcomes measured as binary (the proportion of participants that display the outcome), the efficacy of the test treatment is almost always represented as the absolute difference between the two proportions of outcome in the treatment arms, except for one stroke trial that used risk ratio (RR) as the measure of association. For the 38 continuous outcome variables, they could be reported as the observed true value (n=22, 57.9%), the change from baseline (n=15, 39.5%), or the percentage change from baseline (n=1, 2.6%). The efficacy of the test treatment is measured as the difference between treatment arms, except for 8 diabetes trials that used the ratio of continuous outcome variables between treatment groups as the measure of association. For all of the 9 time-to-event outcomes of interest, hazard ratio was used to compare the efficacy of treatments, and all of the time-to-event outcomes were used within the circulatory disease area (stroke and cardiovascular disease).

Stroke Trials

19 trials were conducted in the stroke area, with the primary outcome variables differing widely between studies. In the trials using time-to-event outcome variables and hazard ratio as the comparison statistic (n=8, 42.1%), the endpoints of interest were generally time to stroke or time to a composite of endpoints including stroke. The NI

margins for hazard ratio had a wide range, between 1.05 and 2. The trial with the largest NI margin had the smallest population sample size (700) and an intermediate study duration (1.5 years), while the trial with the smallest NI margin had the largest population sample size (19000) and the longest study duration (2.5 years). A larger sample size and a longer trial duration will generally improve the power of a study and make it possible to detect a smaller effect difference between treatment groups, thus a smaller NI margin could be detected when the NI margin is decided in advance. The mean of the NI margins expressed as hazard ratio was 1.43, which is kind of high for a serious outcome. It indicated that the test treatment was not inferior to the control treatment if the hazard for stroke (or other outcome) in the test treatment group was at most 43% higher compared with the hazard of stroke in the control group. The 95% confidence interval for the NI margins expressed as hazard ratio ranged between 1.24 and 1.62. Of the 8 stroke trials using binary outcome variables (n=8, 42.1%), only one trial adopted the risk ratio as the comparison statistic and its NI margin was 1.14 with the primary outcome defined as occurrence of death or disability. The other 7 trials used a proportion difference as the comparison statistic, and the NI margins were between 2% and 3% when the outcome was a composite of first stroke or death, which had a relatively low incidence in the population. The NI margin was large (20%) when the outcome was a secondary stroke, which had a high incidence in the study population – children with sickle cell anemia who already had a previous stroke. For the trials focusing on acute stroke remission, NI margins were all 10% regardless of the population size. The remaining 3 trials had continuous outcome variables, and their NI margins were heavily dependent on the particular measure of outcome.

Cardiovascular Trials

20 non-inferiority trials were analyzed in the cardiovascular disease area with 23 identified primary outcome variables. Only one trial applied a time-to-event outcome variable and the NI margin it used was a hazard ratio of 1.3, while its endpoint of interest was a composite of cardiovascular death, nonfatal myocardial infarction or hospitalization due to unstable angina. All analyzed studies on coronary artery disease (n=8) utilized in-stent late lumen loss as the primary endpoint, which was a continuous variable measured as the observed value and the comparison between treatments was shown through the difference of the losses in both groups. Two trials used 0.20mm as the NI margin, three trials used 0.25mm, two trials used 0.32mm, and one trial used 0.11mm. The three trials with target lesion failure as the primary endpoint had similar trial durations of around one year. Their NI margins ranged from 3.5% to 8.6%, and again the trial with the largest sample size had the smallest NI margin, while the trial with the smallest sample size had the largest NI margin. Two trials chose vessel revascularization as the primary endpoint, and the trial with a larger sample size had a smaller NI margin while they shared the same trial duration. Among the five trials using MACE (major adverse cardiac events, defined as adjudicated death, myocardial infarction (MI), or clinically driven target vessel revascularization) as their primary outcome, I again observed that larger trials tend to have smaller NI margins. Sample size in this group ranged from 900 to 2800, with corresponding NI margins ranging from 1.5% to 6.0%. There was one trial with two different NI margins for the MACE outcome, each for a different active control. The remaining cardiovascular trials had unique outcomes and the corresponding NI margins differed significantly due to diverse study design factors.

Pneumonia/Influenza Trials

5 non-inferiority trials on pneumonia treatment were identified, two of them using treatment failure as the primary outcome, and their NI margins were 7% and 8%. One trial used clinical response rate with the NI margin being a 15% absolute proportion difference, one trial used mortality from any cause with the NI margin being a 10% proportion difference, and the final trial used length of hospital stay with the NI margin being 1 day longer. These trials were generally short-term with trial durations between two to four weeks. For three trials on COPD (Chronic Obstructive Pulmonary Disease), an endpoint of FEV1 was used in two of them, and the NI margins were 50mL and 60mL respectively. The third trial used as its outcome the percentage of patients with at least one exacerbation within 6 months after the index exacerbation, and its NI margin was 15%. Among the 13 endpoints reported for NI trials on influenza vaccine, GMTs (geometric mean titers) and seroconversion rate were consistently used to evaluate the efficacy of vaccine. A GMT ratio of 1.5 was selected as the NI margin for all of 7 trials regardless of any other factors, except for one trial with a ratio of 2 as the NI margin. For 6 trials reporting the seroconversion rate, a uniform 10% difference was used for all of them. One additional trial that reported sero-protection rate also utilized a 10% difference as its NI margin.

Diabetes Trials

19 non-inferiority trials were analyzed in diabetes, most of which (n=13, 68.4%) used the change in HbA1c from baseline as the primary outcome and the absolute difference in the changes between the treatment groups as the comparison statistic. The NI margins ranged from 0.3% to 0.5% as absolute values, and the most common NI margin was 0.4% which appeared in 7 out of 13 trials. Other NI margins included: 0.3% used in

two trials, 0.35% used in one trial, 0.45% used in one trial and 0.5% used in two trials. The 2011 FDA guidance on anti-diabetic drugs included the recommendation of a non-inferiority margin of 0.3% HbA1C [19]. Achievement of HbA1c level <=7% was also employed as a binary primary outcome in two trials, and the NI margins were 5% and 7.68% proportion difference in those two trials. A few other outcome variables measuring blood glucose for evaluating diabetes treatments were identified, and their study design features as well as their NI margins are displayed in Table 2.

Overall

After transforming the various scales of comparison statistics into relative scales as discussed earlier, I analyzed a total of 47 NI margins on relative scales out of the original 74 NI trials: 11 NI margins in the stroke area, 15 in the cardiovascular disease area, 16 in the respiratory disease area, and 5 in the diabetes area (Figure 1). The remaining NI trials were excluded or could not be transformed properly because of the following reasons: the trials used hazard ratio as the comparison statistic (9 trials); the trials had the effect size in either group being zero (13 trials); the trials didn't provide any information on either the hypothesized or actual effect sizes of the treatment groups (4 trials); or the trials had a NI margin that exceeded the actual effect size of the active control group with beneficial outcome (2 trials). Because most of the diabetes trials used the change in HbA1c from baseline to the end point as their primary outcome, and the value of change in both groups could sometimes be zero, it was impossible to calculate the ratio of the two changes when denominator may equal to zero, so these trials were excluded. There were 5 transformed NI margins from other trials that used other outcome measures in diabetes. We can see from figure 1 that relative NI margins aggregate between 1 and 2 overall, with the margins

for diabetes being the least variable and the margins for infectious disease being the most variable. The mean of the NI margins for diabetes was 1.24, with a 95% two-sided confidence interval from 1.11 to 1.36. The relative smaller variation in the margins for diabetes is partly due to the well-established guidelines on the selection of NI margins in diabetes trials, which give clear recommendations on the values that researchers could use as the NI margins. The mean of the NI margins for cardiovascular disease was 1.50, with a 95% two-sided confidence interval from 1.38 to 1.62. The mean of the NI margins for stroke was 1.38, with a 95% two-sided confidence interval from 1.23 to 1.53. The mean of the NI margins for infectious disease was 1.38, with a 95% two-sided confidence interval from 1.25 to 1.52. It's surprising that a slightly wider confidence interval and more variability in the NI margins were identified for infectious disease, while in reality the guidance on the selection of the NI margins for infectious disease is much more well-established compared to other disease areas and clear recommendations have been given out in guidance from both FDA and EMA. This phenomenon might result from the diverse severity of the outcome indicators, ranging from the mildest (such as cure or clinical response from pneumonia) to the most severe (such as mortality). The severity of the outcome assessed is an accepted reason to alter the NI margin. I find that the means of the NI margins for different disease domains are similar and the widths of their confidence intervals are nearly the same. Pooling all the trials across different disease areas together, I found a mean of 1.41 and a 95% two-sided confidence interval from 1.33 to 1.48 for the NI margins (Table 3).

Rationales for the Selection of NI Margins

About half of all the trials reported their rationales for selecting a specific non-inferiority margin (Table 4). Among the rationales stated, the most often cited reasons for selecting the NI margin were clinical judgement (18.6%) and guidelines' recommendation (18.6%). Clinical judgement refers to the NI margin that would be considered clinically meaningful or significant to both retain the efficacy of test treatment and prove its non-inferiority to the control. 14.0% of the trials selected the NI margins based on the margins that were used in previous trials, which demonstrated the importance of having reference trials. For NI margins based on the difference between active control and placebo in previous trials (14.0%), various methods were used to decide on the margin, including half of the difference, lower bound of the 95% two-sided confidence interval of the difference, or half of the lower bound [20]. For NI margins based on the difference between test treatment and placebo as well as the difference between active control and placebo in previous trials (11.6%), and those based on the effect sizes of both test treatment and active control in observational studies, the NI margin was selected as the difference between the two differences (for randomized trials) or between the two effect sizes (for observational studies). Other rationales reported included variance of the outcome measurement, expert opinions, and statistical concerns, and were reported in only a few trials.

Discussion

In most of the reports, the rationale behind the chosen NI margin was unclear. *Tsou et al* found that it was unclear how the NI margin was determined for 85% of his reviewed articles [21]. None of the methodological articles referred to as a source for the NI margin contained an exact margin, and none of the previous NI trials referred to contained an

explanation or calculation for the margin. A 2005 review of 332 NI trials identified in extensive literature searches [22] found that only one-fifth of the non-inferiority studies provided an adequate rationale for the selection of the non-inferiority margin. A study of the reporting quality of NI trials published from 2003 to 2004 [23] found that fewer than 20% of the studies reported a clinical consideration and a justification for the margin. In this latter study, the non-inferiority margin was reported in all but a few of the articles. In a cross-sectional study of a random sample of 232 published non-inferiority trials [24], 45.7% reported the method by which the margin was determined and 40.9% of the trials with rationales justified the choice of the NI margin using clinical relevance. The design of a non-inferiority trial requires a trade-off between the possible loss of effectiveness and the benefit of the tested treatment. Another reason for not stating the rationale for the selection of the NI margin also includes the different balancing of priorities between researchers and patients. *Sankoh et al* also found that the margin was missing more frequently in published articles than in unpublished ones, although many of these studies were published more recently after guidance on the transparent reporting of such trials had become available [25]. Journals and authors should be encouraged to improve the reporting on the design and results of NI trials, including information on the rationale of choice for the NI margin and the presentation of the results.

Determining the NI margin implies essentially a trade-off between the possible loss in effectiveness and the perceived advantages of the new treatment from the patient perspectives. Statistical computations should be performed after the NI margin has been determined, after which feasibility can be assessed. After the study is completed, conclusions from non-inferiority trials are based on the judgment of investigators regarding

the NI margin, which may or not correspond to the judgment of patients and clinicians in shared decision making. Thus, constraints regarding logistical and financial burdens should not be included in the reasoning to determine the NI margin, as it should not usually bear on clinical decisions. As with any other study design, reaching an appropriate sample size is important in non-inferiority trials. Even for a treatment that is already significantly non-inferior, the closer the boundary of the confidence interval is to the pre-specified NI margin, the more the decision will be about balancing between the potential loss in effectiveness and the advantages of the new treatment.

No significant association was observed between severity of disease and the range of NI margins. Intuitively, one would prefer to select a smaller margin for trials involving more-serious diseases, given the potential for increased morbidity and mortality rates associated with a less effective drug. If a disease is severe, it is unfavorable to lose much of the effect, even if the costs and the adverse events of the new treatment are lower. Larger margins may cause less concern in trials involving less severe diseases in which treatment failures may not translate into important excess morbidity or mortality. For less severe diseases, the willingness to lose some effect for other benefits might be a reasonable trade off. Therefore, severity of disease could be included in the decision-making process guiding the selection of the NI margin. I did observe that the NI margin was smaller when the primary endpoint was mortality compared with treatment failure, which seems logical as the acceptability of one additional death is usually lower than the acceptability of one additional treatment failure. In a recent experimental study among experts in orthopedic surgery [26], the NI margins reported were also small with median NI margin for risk difference of 1.8% and median NI margin for relative risk of 1.3. The selection of a margin

that is inappropriately large means that the test drug in question may be considerably less effective than the control drug, and it could even be no more effective than placebo. Conversely, with a smaller margin, clinicians have more assurance that a new drug is not inferior to the control agent.

The Division of Anti-Infective Drug Products of the FDA published recommendations [27] on the selection of margins for non-inferiority trials, stating that to establish non-inferiority for a test drug compared with control agent, the statistical analysis of an anti-infective drug trial should employ a 2-tailed 95% CI around the difference in outcomes. One approach to setting the lower bound of the 95% CI would be to base this limit on the success rate achieved in the trial. The recommendation was for use of 10%, 15%, and 20% deltas in trials with success rates of 90%, 80%–89%, and 79%, respectively. Since then, the majority of comparative trials is this disease area have used this so-called "step function" for selecting their margins [28]. The ICH thus recommends [29] that investigators base the selection of margins first on the magnitude of benefit of any therapy over placebo or no treatment in randomized controlled trials involving the disease in question. If the magnitude of such an effect is known and is sufficiently large, then one can select the non-inferiority margin based on the acceptable potential loss of efficacy over existing therapies. The determination of the benefit of therapy over placebo should come from actual data, but the determination of what constitutes an acceptable potential loss of efficacy, relative to existing therapies, is based on clinical judgment. The ICH guidelines implicitly recommend disease-specific selection of margins. However, a disease-specific approach raises the question of how to select margins for clinical trials in which the exact magnitude of the benefit of therapy may be unclear. The questions also arise of what

elements of clinical judgment are important in selecting NI margins and how these elements should be applied for diseases of varying severity. With use of smaller margins, the sample size for the trial increases. This may have implications on the ability to perform a trial above and beyond the time and monetary considerations. For a relatively rare, serious disease with a low cure rate, the number of patients needed to perform the trial with a smaller margin could even exceed the number of patients who have that disease diagnosed annually in the United States.

Vaccine trials are unique among infectious disease trials because most vaccines are highly effective and work to prevent disease. For preventive interventions, a small NI margin might be preferable because losing a large part of the effect on efficacy or immunogenicity is undesirable. *Donken et al* found that many vaccine studies used an NI margin of 10% for the difference (66%) and 0.67/1.5 or 0.5/2.0 for the GMT ratio (94%), as suggested in guidelines [30]. Although in most cases the margins used were in line with the recommended margins, only 6% of the articles stated that they followed either EMA or FDA guidelines. The differences and the variability in NI margins used might be because of the lack of clear and explicit guidelines on which NI margin to use for vaccine trials. It is also questionable whether all vaccine trials should use the same NI margin. Examples of possible factors to consider are the severity of disease, side effects, potential benefits of the new vaccine, the aim or study objective and the quantity and quality of the immune response. It is important to choose the NI margin in a way that coverage or percentage protected in the total population will be sufficient to prevent the spread of the disease.

Determination of the NI margin could be simplified by better and more explicit guidelines from the regulatory authorities. An improved approach might be to employ the

FDA method of using a preserved fraction if the study design allows this. A preserved fraction of 50% has become common practice in non-inferiority trials (e.g. cardiovascular, and irreversible morbidity), while higher fractions have been used in other disease areas (e.g. 90% preserved fraction in antibiotics). However, problems also arise from this approach. For example, consider an attempt to prove that drug B is not inferior to drug A (active comparator), and the effect size of drug A from previous placebo-controlled trials is 60%, which indicates that drug A will show effectiveness in 60% of the patients. By using a preserved fraction of 50%, we have the NI margin as 30%, meaning that an effect size as low as 30% for drug B is enough to conclude that drug B is not inferior to drug A. Then, consider if drug B is widely used while drug A is withdrawn from the market and no longer available. If we now test a new NI trial trying to prove that drug C is not inferior to drug B, and based on the effect size of drug B being 30% and a preserved fraction of 50%, an effect size as low as 15% for drug C is enough to prove that drug C is not inferior to drug B. If this situation continues, we would get less and less effective tested treatments while the non-inferiority still holds without any obvious violations in the selection of the NI margins. To avoid this "chain reaction" and prevent the situation that the effects of the tested treatments are shrinking all the way down, one possibility might be to establish a starting or fixed value for the NI margin, and then this NI margin could be adapted based on the specific characteristics of that NI trial. For example, we could set the NI margin as 10% for all NI trials in this disease area, and the NI margin in a specific trial could be adjusted within the range of 8% to 12%, considering factors including the seriousness of the outcome, risk-benefit profile of the active comparator, or whether the effect of the active comparator has diminished over time. Another possibility to avoid this "chain

reaction" would be to use the same active comparator for all tested treatments, for example, to use drug A as the active comparator for the evaluations of both drug B and drug C.

Limitations

The primary limitation of my analyses is the potential for publication bias. The published trials included in this study may not be representative of all NI trials published in the literature. Another limitation of my study is the relatively small sample size. It is possible that a larger study, including more trials, might identify predictors associated with the selection of non-inferiority margins and confidence intervals or P-values. Finally, the process of determining the non-inferiority margin was described only in general terms in most published articles. Thus, it was difficult to analyze the reasoning behind the selected non-inferiority margin.

In summary, one should select a margin for a NI clinical trial that allows a high degree of certainty that the drug is better than placebo and ensures that any potential loss in efficacy compared to an active control is clinically acceptable. With use of a smaller margin, one achieves a higher degree of certainty that the efficacies of the drug being studied in the trial is indeed non-inferior. However, one must balance the desire for more certainty against the current state of knowledge about a given disease, clinical judgment, and the practicalities of performing clinical trials including the required sample size. Audiences have been introduced to the concepts and terminology of this study design as well as to the more detailed issues of conduct and analysis. Regulatory agencies have provided guidance and the FDA issued a draft guidance to highlight their current recommendations on acceptable approaches to the design and analysis of NI trials. Although there is not an established 'optimal' approach or design, most parties in the

clinical trial community agree that a successful NI trial requires the key elements below: appropriate active control, appropriate NI margin, overall assay sensitivity, and limited cross-overs.

In practice, the selection of non-inferiority margins should be based on both statistical justification and clinical judgment. More research is necessary on how to integrate statistical justification and clinical judgment in the development of non-inferiority margin. In the meantime, it is essential that regulators are aware of the difficulties faced by applicants, and scientific dialogue between both parties can support the regulators in improving guidance on non-inferiority trials.

Conclusion

Generally, it appears arbitrary when researchers select the NI margins for non-inferiority trials. It appears that they select the NI margins without solid rationales, and the NI margins are more like a product under the manipulation in the purpose of making the trial feasible to conduct and statistically significant. Difficulties still exist in selecting the appropriate NI margin of non-inferiority trials. Straightforward and harmonized guidance on the selection of non-inferiority margin is needed. It is unlikely that regulatory guidelines can cover all therapeutic areas, either as one general guideline or special sections on non-inferiority trials in disease-specific guidelines, however, they could be used as an opportunity for tailored advice in many cases.

Bibliography

[1] Kim, M. Y., & Xue, X. (2004). Likelihood ratio and a Bayesian approach were superior to standard noninferiority analysis when the noninferiority margin varied with the control event rate. *Journal of clinical epidemiology*, *57*(12), 1253-1261.

[2] Dekkers, O. M., Cevallos, M., Bührer, J., Poncet, A., Rau, S. A., Perneger, T. V., & Egger, M. (2015). Comparison of noninferiority margins reported in protocols and publications showed incomplete and inconsistent reporting. *Journal of clinical epidemiology*, *68*(5), 510-517.

[3] Gayet-Ageron, A., Agoritsas, T., Rudaz, S., Courvoisier, D., & Perneger, T. (2015). The choice of the noninferiority margin in clinical trials was driven by baseline risk, type of primary outcome, and benefits of new treatment. *Journal of clinical epidemiology*, *68*(10), 1144-1151.

[4] Cleophas, T. J., & Zwinderman, A. H. (2012). A Novel Approach to Noninferiority Testing: Documented Proof Rather Than Arbitrary Margins. *American journal of therapeutics*, *19*(2), 101-107.

[5] ICH E10 2000. Choice of control group in clinical trials. Available at URL: *http://www.ich.org/LOB/media/MEDIA486.pdf* [14 May2011].

[6] US Food and Drug Administration. (2010). Guidance for industry non-inferiority clinical trials. *US Department of Health and Human Services and US Food and Drug Administration, Washington, DC*.

[7] Koopmeiners, J. S., & Hobbs, B. P. (2016). Detecting and accounting for violations of the constancy assumption in non-inferiority clinical trials. *Statistical Methods in Medical Research*, 0962280216665418.

[8] Prins, M. H., & Lensing, A. W. (2013). Derivation of the non-inferiority margin for the evaluation of direct oral anticoagulants in the treatment of venous thromboembolism. *Thrombosis journal*, *11*(1), 13.

[9] Tsong, Y., Zhang, J., & Levenson, M. (2007). Choice of δ Noninferiority Margin and Dependency of the Noninferiority Trials∗. *Journal of biopharmaceutical statistics*, *17*(2), 279-288.

[10] Powers, J. H., Ross, D. B., Brittain, E., Albrecht, R., & Goldberger, M. J. (2002). The United States Food and Drug Administration and noninferiority margins in clinical trials of antimicrobial agents. *Clinical infectious diseases*, *34*(6), 879-881.

[11] Committee for Medicinal Products for Human Use. (2006). Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. *Statistics in medicine*, *25*(10), 1628.

[12] Durkalski, V., Silbergleit, R., & Lowenstein, D. (2011). Challenges in the design and analysis of non-inferiority trials: a case study. *Clinical Trials*, *8*(5), 601-608.

[13] Gupta, R., Gupta, H., & Banker, M. (2016). The impact of an inappropriate non-inferiority margin in a non-inferiority trial. *Human Reproduction*, *31*(12), 2892-2893.

[14] Hou, Y., Wu, X. Y., & Li, K. (2009). Issues on the selection of non-inferiority margin in clinical trials. *Chinese medical journal*, *122*(4), 466-470.

[15] Gerlinger, C., & Schmelter, T. (2011). Determining the non-inferiority margin for patient reported outcomes. *Pharmaceutical statistics*, *10*(5), 410-413.

[16] Colagiuri R, Dain K, Moylan J. The global response to diabetes: action or apathy? *MJA*. 2014;201(10):581–3.

[17] International Diabetes Federation. IDF Diabetes Atlas. In: Seventh Edition edn. *Brussels: International Diabetes Federation;* 2015.

[18] Matilde Sanchez, M., & Chen, X. (2006). Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. *Statistics in medicine*, *25*(7), 1169-1181.

[19] Wangge, G., Putzeist, M., Knol, M. J., Klungel, O. H., Gispen-De Wied, C. C., de Boer, A., ... & Mantel-Teeuwisse, A. K. (2013). Regulatory scientific advice on non-inferiority drug trials. *PloS one*, *8*(9), e74818.

[20] Julious, S. A. (2011). The ABC of non-inferiority margin setting from indirect comparisons. *Pharmaceutical statistics*, *10*(5), 448-453.

[21] Tsou, H. H., Hsiao, C. F., Chow, S. C., Yue, L., Xu, Y., & Lee, S. (2007). Mixed noninferiority margin and statistical tests in active controlled trials. *Journal of biopharmaceutical statistics*, *17*(2), 339-357.

[22] Ng, T. H. (2008). Noninferiority hypotheses and choice of noninferiority margin. *Statistics in medicine*, *27*(26), 5392-5406.

[23] Nie, L., & Soon, G. (2010). A covariate-adjustment regression model approach to noninferiority margin definition. *Statistics in medicine*, *29*(10), 1107-1113.

[24] Spellberg, B., Talbot, G. H., Boucher, H. W., Bradley, J. S., Gilbert, D., Scheld, W. M., ... & Antimicrobial Availability Task Force of the Infectious Diseases Society of America. (2009). Antimicrobial agents for complicated skin and skin-structure infections: justification of noninferiority margins in the absence of placebo-controlled trials. *Clinical infectious diseases*, *49*(3), 383-391.

[25] Sankoh, A. J. (2008). A note on the conservativeness of the confidence interval approach for the selection of non-inferiority margin in the two-arm active-control trial. *Statistics in medicine*, *27*(19), 3732-3742.

[26] Wang, Y., Harigaya, Y., Cavaillé-Coll, M., Colangelo, P., & Reynolds, K. S. (2015). Justification of noninferiority margin: Methodology considerations in an exposure–response analysis. *Clinical Pharmacology & Therapeutics*, *97*(4), 404-410.

[27] Gosho, M. (2015). Non-inferiority margins employed in clinical trials in Japan. *Journal of clinical pharmacy and therapeutics*, *40*(3), 289-298.

[28] Althunian, T. A., Boer, A., Groenwold, R. H., & Klungel, O. H. (2017). Defining the non-inferiority margin and analyzing non-inferiority: An overview. *British Journal of Clinical Pharmacology*.

[29] Chow, S. C., & Shao, J. (2006). On non-inferiority margin and statistical tests in active control trials. *Statistics in medicine*, *25*(7), 1101-1113.

[30] Donken, R., De Melker, H. E., Rots, N. Y., Berbers, G., & Knol, M. J. (2015). Comparing vaccines: A systematic review of the use of the non-inferiority margin in vaccine trials. *Vaccine*, *33*(12), 1426-1432.

Tables

Table 1. Basic characteristics of the included non-inferiority trials (N=74)

Characteristics	No. of trials	%
Randomized		
Yes	73	98.6
No	1	1.4
Masking		
Open-label	24	32.4
Subject-mask	7	9.5
Observer-mask	15	20.3
Double-mask	18	24.3
No	4	5.4
Unknown	6	8.1
Study phase		
II/IIb	2	2.7
III	9	12.2
IV	4	5.4
Unknown	59	79.7
No. of treatment arms		
2	66	89.2
3	8	10.8
Use of placebo control		
Yes	3	4.0
No	71	96.0
Participants		
Children	7	9.5
Adults	53	71.6
Unknown	14	18.9
Regions		
USA	14	18.9
Non-USA	26	35.1
Multinational	34	46.0
Primary analysis population		
ITT[a]	36	48.6
PP[b]	15	20.3
Both	17	23.0
Unknown	6	8.1
Significance level (two-sided)		
0.05 (95% CI[c]**)**	70	94.6
0.1 (90% CI)	2	2.7
Unknown	2	2.7

[a] ITT, intention-to-treat.

[b] PP, per-protocol.

[c] CI, confidence interval.

Table 2. Design of the included non-inferiority trials and their non-inferiority margins

Disease domain[a]	No. of trials	Trial duration[b]	Sample size	Primary endpoint					
				Definition[c]	Type[d]	Unit	Scale[e]	CS[f]	NI margin
Stroke									
	2	2.5Y; 2Y	2700; 18000	Occurrence of stroke	TTE	-	-	HR	1.33; 1.46
	3	1.5Y; 1Y; 2.5Y	700; 1600; 19000	Composite of stroke, myocardial infarction, or vascular events	TTE	-	-	HR	2; 1.25; 1.05
	3	8M; 10M; 2Y	3700; 4600; 15000	Composite of stroke or systemic embolism	TTE			HR	1.38; 1.5; 1.46
	3	30D; 30D; 5Y	500; 1200; 1500	Stroke or death	Bin	%	-	Dif	2; 2.5; 3
	1	2.5Y		Occurrence of secondary stroke	Bin	%	-	Dif	20
	1	4W	120	P2Y12 percentage inhibition	Con	% points	CB	Dif	9
	1	4W	200	Global CBF change	Con	%	CB	Dif	8.6
	1	2Y	120	TCD velocity	Con	cm/s	Obs	Dif	15
	1	90D	450	Modified Rankin Scale score of 0-2	Bin	%	-	Dif	10
	1	90D	200	Successful recanalization	Bin	%	-	Dif	10
	1	90D	110	TIMI scale 2 or 3 flow	Bin	%	-	Dif	10
	1	90D	3300	Death or disability	Bin	%	-	RR	1.14

Table 2 (continued)

Disease domain[a]	No. of trials	Trial duration[a]	Sample size	Primary endpoint				CS[e]	NI margin
				Definition[b]	Type[c]	Unit	Scale[d]		
CVD									
	1	3Y	15000	Composite of cardiovascular death, myocardial infarction, or hospitalization	TTE	-	-	HR	1.3
	8	6M; 6M; 6M; 9M; 13M; 9M; 9M; 8M	180; 250; 300; 350; 400; 450; 1100; 1700	In-stent late lumen loss	Con	mm	Obs	Dif	0.25; 0.32; 0.20; 0.25; 0.20; 0.11; 0.25; 0.32
	3	13M; 12M; 12M	400; 2000; 2300	Target lesion failure	Bin	%	-	Dif	8.6; 4.5; 3.5
	2	9M; 9M	1100; 1800	Target vessel revascularization	Bin	%	-	Dif	3.6; 3
	5	12M; 8M; 9M; 9M; 12M	900; 1700; 1700; 2800; 5700	MACE	Bin	%	-	Dif	6; 5; 4; 1.5; 2 and 2.5
	1	12M	130	Freedom from MACE	Bin	%	-	Dif	10
	1	12M	3100	NACCE	Bin	%	-	Dif	2.7
	1	30D	3750	Composite of cardiac deaths, myocardial infarction, stroke, or major bleeding	Bin	%	-	Dif	0.75
	1	30D	70	Inhibition of platelet aggregation (%IPA)	Bin	%	-	Dif	5

Table 2 (continued)

Disease domain[a]	No. of trials	Trial duration[a]	Sample size	Primary endpoint				CS[e]	NI margin
				Definition[b]	Type[c]	Unit	Scale[d]		
Diabetes									
	1	1D	375	Glycemic control measured by mean daily BG concentration	Con	mg/dL	Obs	Dif	18
	1	26W	180	Baseline adjusted HbA1c	Con	mmol /mol	Obs	Dif	5.5
	13	24W; 24W; 24W; 6M; 6M; 20W; 28W; 24W; 24W; 12W; 6M; 26W; 78W	120; 150; 160; 160; 186; 250; 270; 300; 375; 490; 500; 520; 800	Change in HbA1c	Con	%	CB	Dif	0.30; 0.50; 0.40; 0.40; 0.40; 0.40; 0.45; 0.50; 0.30; 0.35; 0.40; 0.40; 0.40
	2	24W; 24W	320; 970	Achievement of an HbA1c level of <=7%	Bin	%	-	Dif	5; 7.68
	1	6W	80	Blood fructosamine	Con	%	PB	Dif	20
	1	32W	80	Coefficient of variation of FBG	Con	-	Obs	Ratio	1.25
COPD									
	2	12W; 26W	657; 676	FEV1	Con	mL	Obs	Dif	50; 60
	1	6M	183	Patients with at least one exacerbation	Bin	%	-	Dif	15

Table 2 (continued)

Disease domain[a]	No. of trials	Trial duration[a]	Sample size	Primary endpoint					
				Definition[b]	Type[c]	Unit	Scale[d]	CS[e]	NI margin
Pneumonia									
	2	14D; 90D	520; 580	Treatment failure	Bin	%	-	Dif	7; 8
	1	7-14D	300	Clinical responses at test-of-cure	Bin	%	-	Dif	15
	1	28D	496	Mortality from any cause	Bin	%	-	Dif	10
	1		130	Length of hospital stay	Con	Days	Obs	Dif	1
Influenza vaccine									
	7	30 & 180D; 21D; 28D; 30D; 28D; 21D; 21-28D	187; 320; 1250; 1300; 1670; 1680; 1850	GMTs	Con	-	Obs	Ratio	1.5; 2; 1.5; 1.5; 1.5; 1.5; 1.5
	5	28D; 30D; 28D; 21D; 21-28D	1250; 1300; 1670; 1680; 1850	Seroconversion	Bin	%	-	Dif	10; 10; 10; 10; 10;
	1	30D	300	Sero-protection	Bin	%	-	Dif	10

[a] CVD, cardiovascular disease; COPD, chronic obstructive pulmonary disease.

[b] Y, years; M, months; W, weeks; D, days.

[c] CBF, cerebral blood flow; TCD, transcranial Doppler; TIMI, Thrombolysis in Myocardial Ischemia; MACE, major adverse cardiac events, defined as adjudicated death, myocardial infarction (MI), or clinically driven target vessel revascularization (TVR); NACCE, net adverse clinical and cerebral events, a composite of all-cause death, myocardial infarction [MI], stroke, or major bleeding; FEV1, trough forced expiratory volume in 1 second; GMTs, geometric mean titers; BG, blood glucose; FBG, fasting blood glucose.

[d] TTE, time to event; Bin, binary; Con, continuous.

[e] Obs, observed value; CB, change from baseline; PB, percentage change from baseline.

[f] CS, comparison statistic; HR, hazard ratio; Dif, difference; RR, risk ratio.

Table 3. Non-inferiority margins on relative scales by disease domain

Disease domain	No. of margins	Mean	95% CI[a]	Range
Stroke	11	1.38	1.23 – 1.53	1.1 – 1.88
Cardiovascular disease	15	1.50	1.38 – 1.62	1.06 – 1.96
Infectious disease	16	1.38	1.25 – 1.52	1.04 – 2.00
Diabetes	5	1.24	1.11 – 1.36	1.08 – 1.43
Total	47	1.41	1.33 – 1.48	1.04 – 2.00

[a] CI, confidence interval.

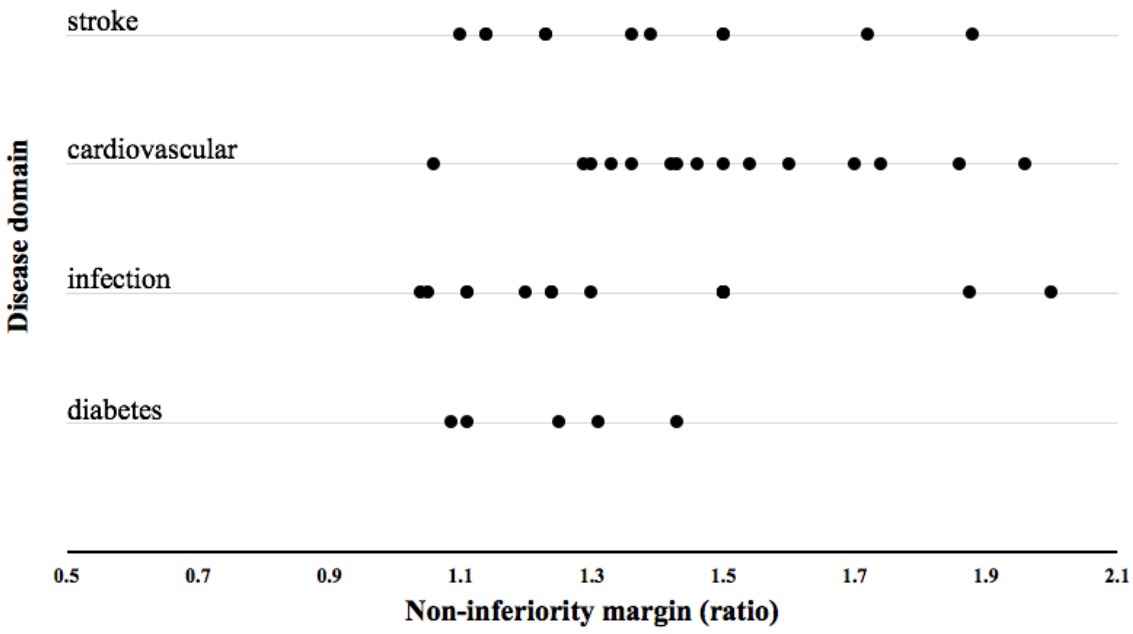
Table 4. Rationales for the selection of non-inferiority margins in the included NI trials

Rationales	No. of trials[a]	%
Based on the effect size of active control over placebo in previous trials	6	14.0
Based on both the effect size of test treatment over placebo and the effect size of active control over placebo in previous trials	5	11.6
Based on the effect size of active control in previous non-trial studies	1	2.3
Based on both the effect size of test treatment and the effect size of active control in previous non-trial studies	5	11.6
Based on the NI margin used in previous NI trials	6	14.0
Based on the variation of the primary outcome measurement	2	4.6
Clinically meaningful/significant judgement	8	18.6
Guidelines/recommendations	8	18.6
Experts opinions	1	2.3
Statistically appropriate/feasible	1	2.3
Total	43	100

[a] Number of trials is duplicate counting, in which one NI trial may contribute to multiple rationales for the selection of NI margin.

Figures

Figure 1: Non-inferiority margins on relative scales for included NI trials by disease domain



CV - Ling Yang

Tel: +1 443-743-6091 • E-mail: lyang52@jhu.edu • Address: 1 Cross Keys Road, Apt. E, Baltimore, MD 21210

EDUCATION

Bloomberg School of Public Health **Johns Hopkins University** Baltimore, MD

Master of Science in Epidemiology (Cumulative GPA: 3.91/4.00) Track: Clinical trials and evidence synthesis Sep. 2015 – May. 2017

Certificate in Pharmacoepidemiology and Drug Safety

Relevant Courses: Epidemiologic Methods; Biostatistics; Pharmacoepi. Methods; Systematic Reviews & Meta-analysis; Analysis of Longitudinal Data; Survival Analysis; Clinical Trials Mgmt.; Public Health Surveillance

School of Public Health **Peking University Health Science Center** Beijing, China

Bachelor of Medicine (Cumulative GPA: 3.80/4.00, Ranking: 1/85) Sep. 2010 – Jul. 2015

National School of Development **Peking University** Beijing, China

Bachelor of Economics (Cumulative GPA: 3.80/4.00) Sep. 2011 – Jul. 2015

WORK EXPERIENCE

Research Assistant – Johns Hopkins School of Public Health (JHSPH) May. 2016 – Present

- Conducted longitudinal paired-visit analysis to examine Pap/HPV concordance by age and other factors including follow-up types
- Implemented extensive literature reviews on the existence and extent of misclassification between stillbirth and neonatal death
- Collected and abstracted data from Demographic & Health Surveys (DHS) and Verbal Autopsy & Social Autopsy (VASA) Studies
- Performed cross-sectional analysis to describe the misclassification and fitted logistic models to examine its risk factors
- Initiated a systematic review on the coverage of four children vaccines (measles, influenza, pneumonia & rotavirus) by time
- Mapped rates of child sexual abuse at jurisdiction level in the US and explored its spatial/geographical variation using ArcGIS
- Involved in study coordination, protocol development, data management & analysis and drafting of materials for publication

Intern – Beijing SHIJITAN Hospital / 302 Military Hospital of China Sep. 2013– Sep. 2014

- Collected and completed medical records, performed medical procedures including physical diagnosis in various departments
- Aided doctors in making prescriptions & changing dressings in inpatient department and guided patients to finish X-ray & MRI
- Performed surgical operations (lipomyoma, inguinal hernia, fibroid & cancer) as an assistant, tracked and evaluated prognosis
- Assisted with data collection and entry for Army Respiratory Diseases Investigation project relating to the symptoms and signs
- Processed epidemiological investigation of Ebola hemorrhagic fever and malaria through patients query and interview

Teaching Assistant / Section Instructor – JHSPH and Johns Hopkins Homewood Campus Jul. 2016 – Present

- Courses: *Principles of Epidemiology, Fundamentals of Epidemiology, Epidemiologic Methods 3, Intro to Epidemiology, Stata*
- Led discussion sessions for 25 students, arranged them in small groups, presented the slides and answered their questions
- Taught in labs with other TA and a faculty instructor, held office hours, maintained course website, communicated with students
- Developed course materials in conjunction with instructors, drafted exam questions and graded exams & assignments

RESEARCH EXPERIENCE

Impact of Voluntary Medical Male Circumcision (VMMC) on Women's Health – Johns Hopkins Program for International Education in Gynecology and Obstetrics (Jhpiego) and U.S. CDC Jan. – May. 2016

- Systematic review on the impact of VMMC on women's health regarding HIV and other sexually transmitted infections (STIs)
- Trainings on systematic review conduction procedure, PRISMA checklist and quality assessment scales on all study types
- Title/abstract screening, full-text screening, data abstraction from article contents using Access and discrepancies harmonization
- Data collection regarding the quality of studies, quality grading and assessment based on scales corresponding to the study type
- Search update and repeating previous phases for the new findings, manuscript revision and preparation for article publication
- Article *Impacts of Male Circumcision on Women's Biomedical Health Outcomes: a systematic review* is **ready to publish**

Effects of Occupational Exposure to Chromate on Worker's Health – Peking University Health Science Center Sep. 2014– Jul. 2015

- Facilitated field investigation in chemical factory, collected workers' information through questionnaire and physical examination
- Performed cytokinesis-blocked micronucleus experiment and conducted statistics on micronucleus shown under microscope
- Managed and stored materials, analyzed data (description and comparison) using SPSS, drafted manuscript, presented results
- Reviewed systematically on the reporting quality assessment of network meta-analysis based on PRISMA and AMSTAR checklists
- Cultivated a number of research skills including project design, data collection through questionnaire administration and data entry

PROFESSIONAL DEVELOPMENT

Computer Skills: Biostatistics computing software (STATA, SAS, SPSS, R); Microsoft Office suite (Word, Excel, PowerPoint, Access); ArcGIS, Google Earth Pro; EpiData, EndNote, RefWorks

Training: Human Subjects Research, Collaborative Institutional Training Initiative (CITI) Program; Good Clinical Practices (GCP)

PUBLICATIONS

T.J. Wang, **Ling Yang**, G. Jia. *Effects of Vitamin C on the Genotoxicity of Hexavalent Chromium*. Toxicology, 2016 Feb, 30(1): 87-90.

Ling Yang, W. Zhu. *Early Intervention and Prevention of Postnatal Depression*. Chinese Journal of Public Health, 2014 Mar, 30: 114-116.